We have the following database (<https://www.uv.es/mperea/readingskill.sav>) which contains data on reading ability through the PROLEC test in high school students, as well as the average number of words learned incidentally averaged across several classes. We also have the answers (on a Likert scale of 1-5) to an extroversion test: for the first 2 questions, the higher the score, the higher the level of extroversion (e.g., "my friends consider me a social person"; "I prefer to go to a party rather than to read a book") and the third is the inverse (the higher score, the lower the extroversion; e.g., "I consider myself a quiet person").

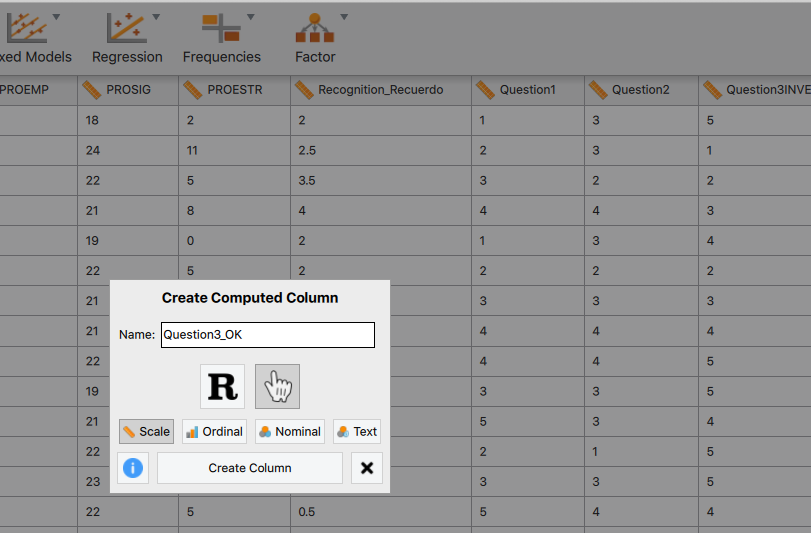
*1. Are there any differences in extroversion between boys and girls?*

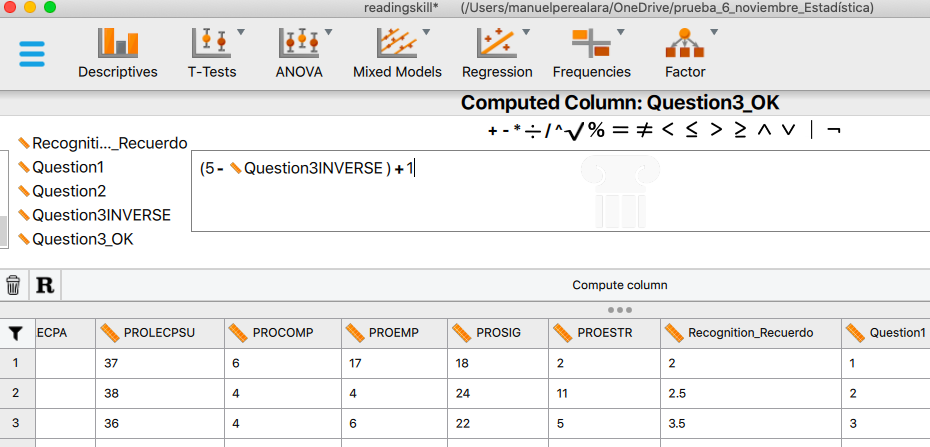
*(To have the overall “extroversion”, you have to add up all 3 questions; the third question must be coded in the right order: the higher the score, the higher the level of extroversion) Remember that the Likert scale questions should be considered as quantitative.*

First, we must have the 3 Likert-type questions as quantitative ("scale") and have question 3 in the right order.

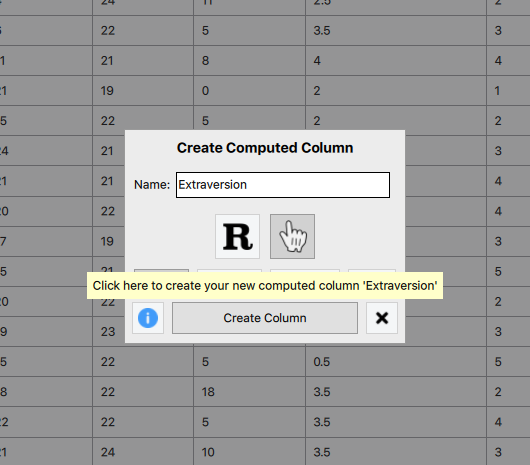
The formula is 5 - REVERSE\_ORDER\_QUESTION + 1

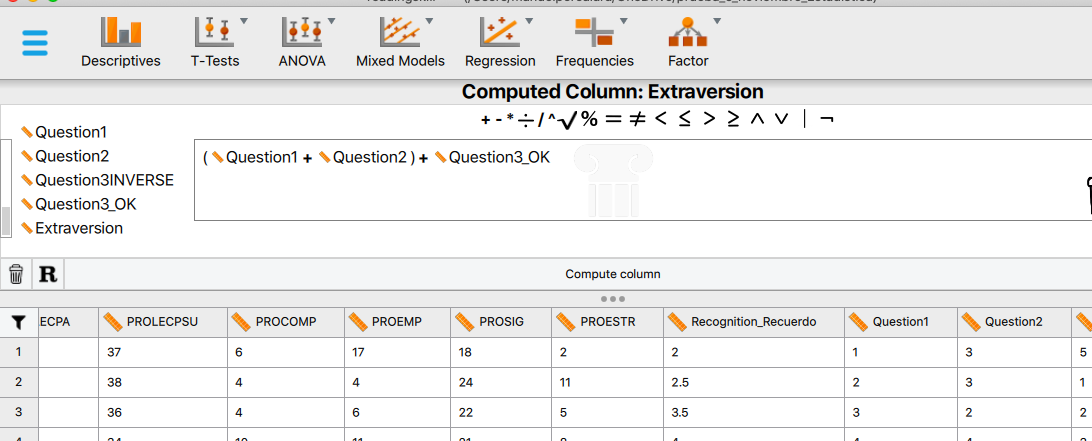
(5 is the maximum value of the Likert scale; if it were a Likert scale of 1-10 then the number would be 10 instead of 5, logically)



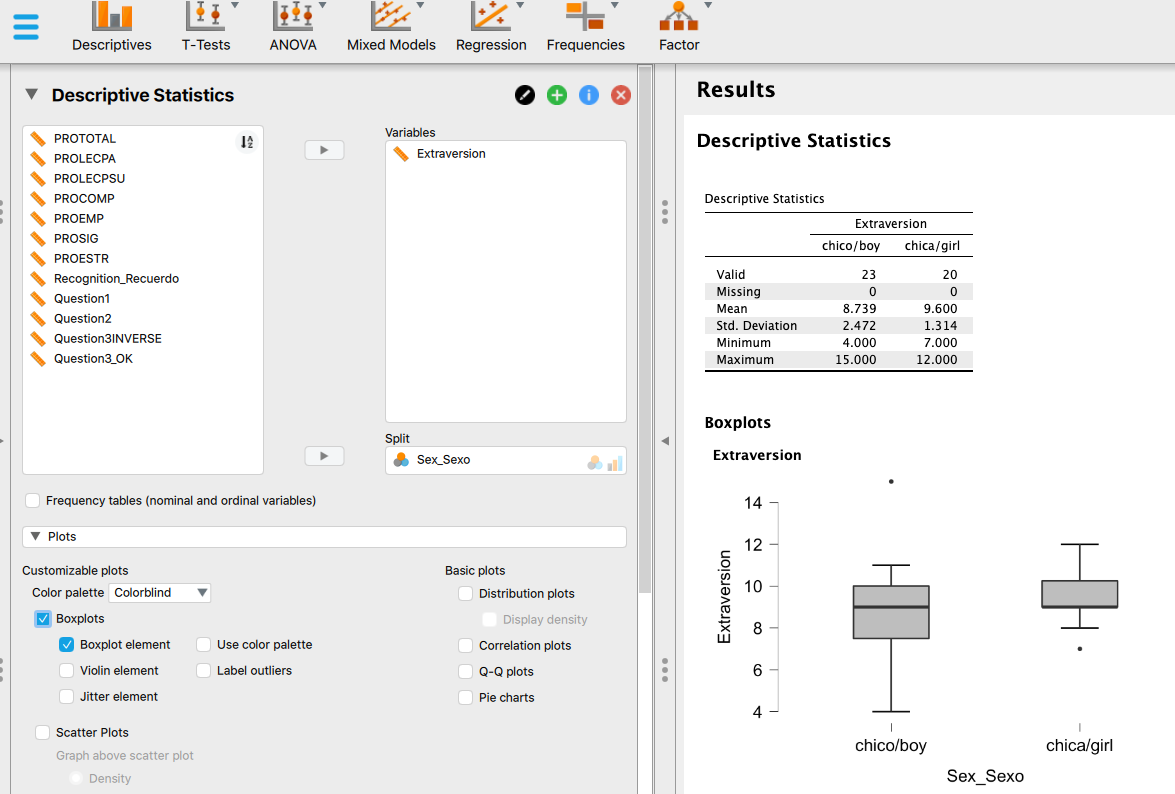


And now we have to add the 3 questions in the new variable "extraversion”





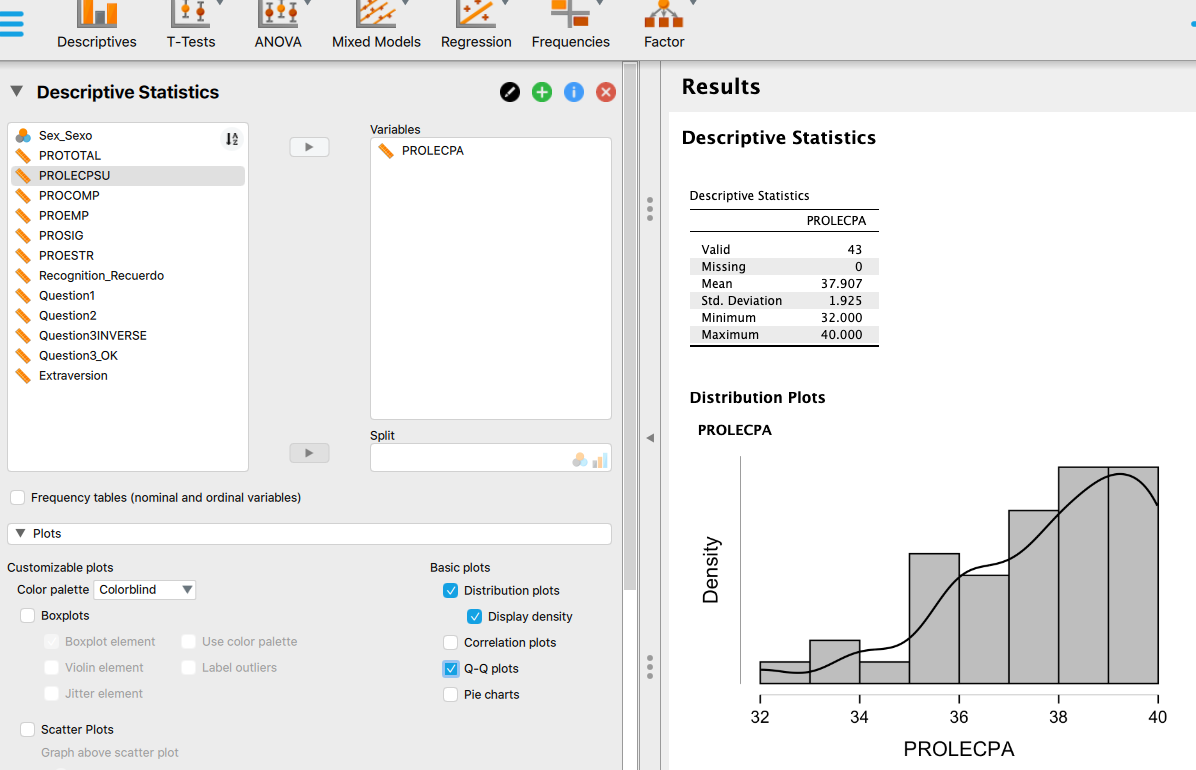
Now, you can have a look at the box diagram and the boys' and girls' averages. Extraversion is the dependent variable and sex is the independent variable.



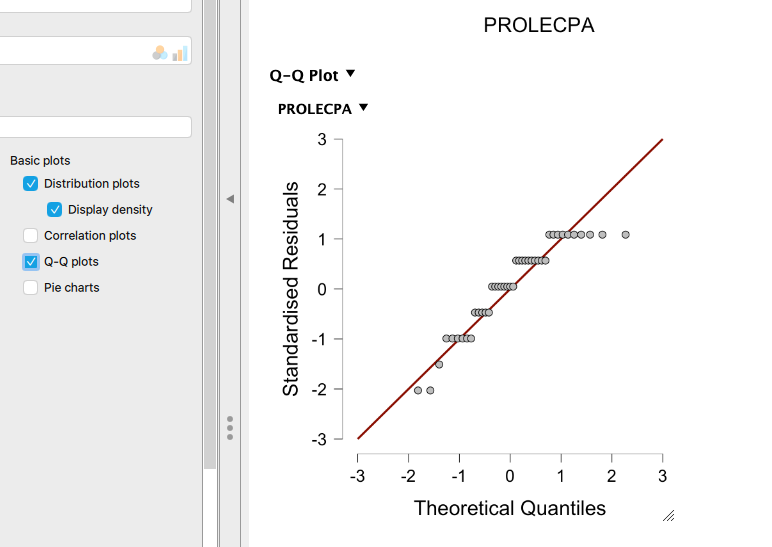
In the sample, girls show somewhat higher levels of extroversion than boys (M = 9.6 vs. 8.7, respectively). In addition, there seems to be less variability in the group of girls than in that of boys (SD = 1.3 vs. 2.5, respectively).

*2. Overall, does the PROLECPA test (word reading measurement in the PROLEC test) follow approximately a normal distribution? What is the asymmetry/skewness? What about kurtosis?*

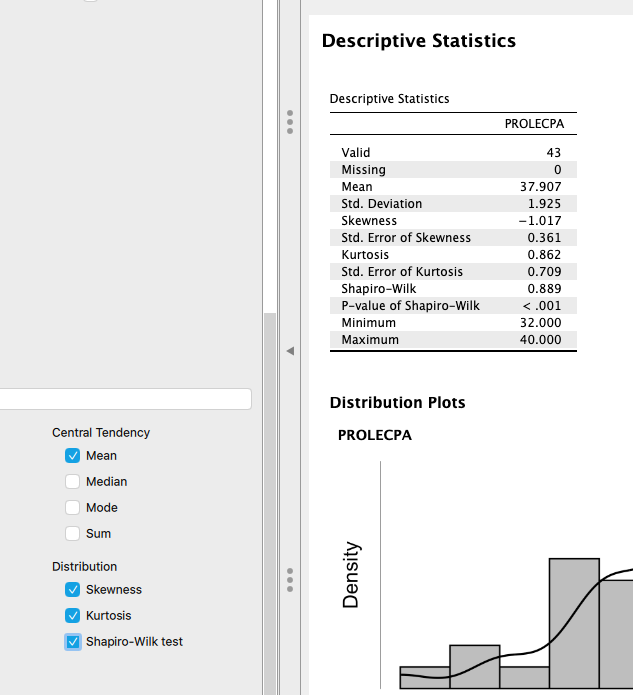
The best strategy to begin with is to look at the histogram (smoothed):



It shows a clear negative asymmetry, so the distribution does not seem to be normal. Also, in the Q-Q Plot, the points do not seem to be close to the line either.



If we want to obtain statistics (remember that the normal distribution is symmetric and its shape is "mesocurtic"), we can calculate the index of asymmetry and the index of kurtosis. And also apply the Shapiro-Wilks test (if the probability value p obtained is less than 0.05 then it is very unlikely that the data follow the normal distribution).

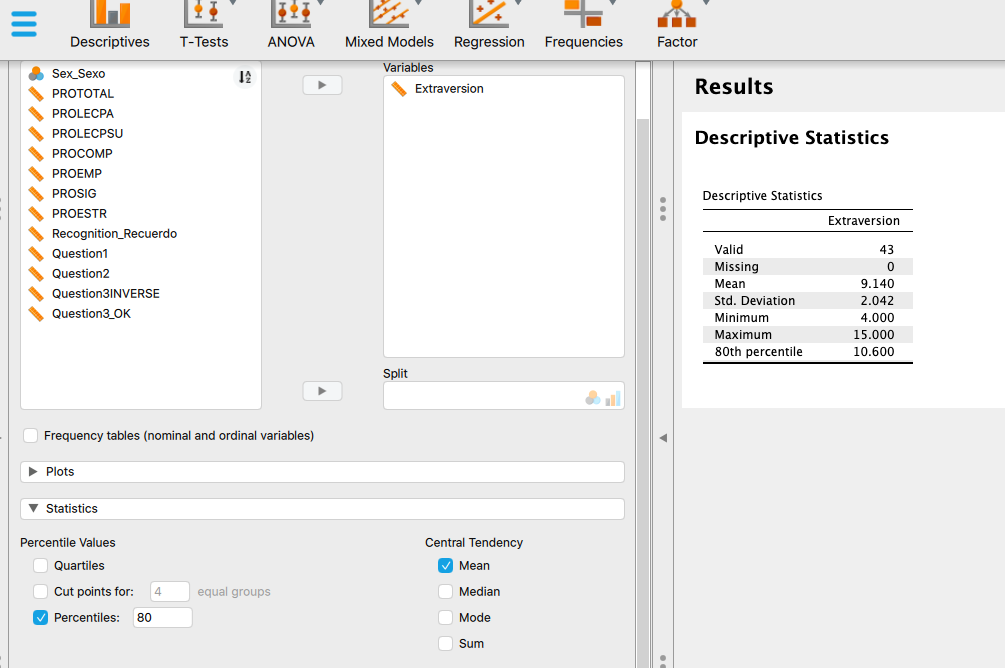


It can be seen that the asymmetry index is negative, as expected, -1.017, and that the distribution looks like leptocurtic (kurtosis index = 0.86). Further, the Shapiro-Wilk test shows a very small p value....

Therefore, both the statistics and the plots clearly show that the PROLECPA distribution does not follow the normal distribution.

*3. If we want to select the 20% most extraverted people for a communication course, what cut-off point will we choose?*

The answer is the 80th percentile, which is the score that leaves above 20% of data.

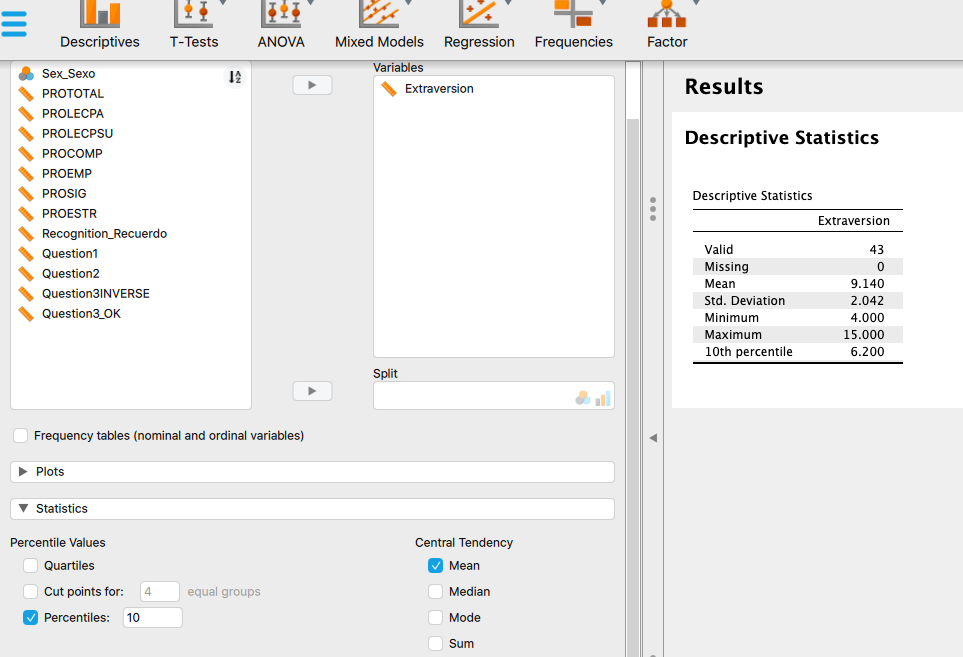


This value is 10.6

*4. If we want to select the 10% most introverted people for a social skills course, what cut-off point will we choose?*

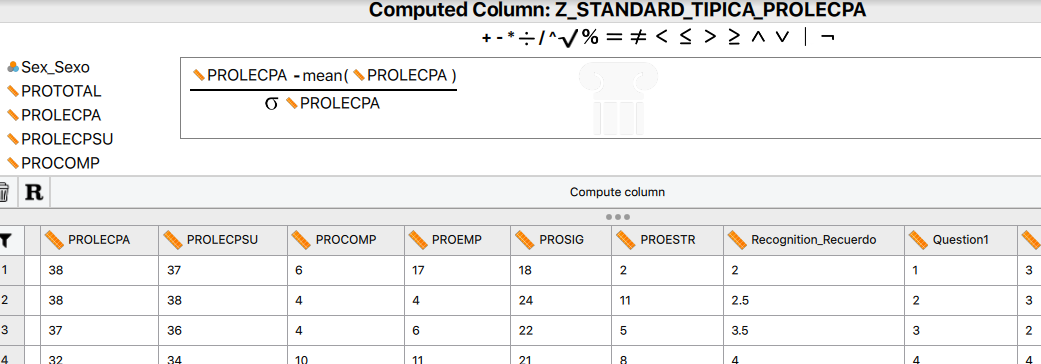
The 10th percentile, which is the score that leaves below the 10% of data.

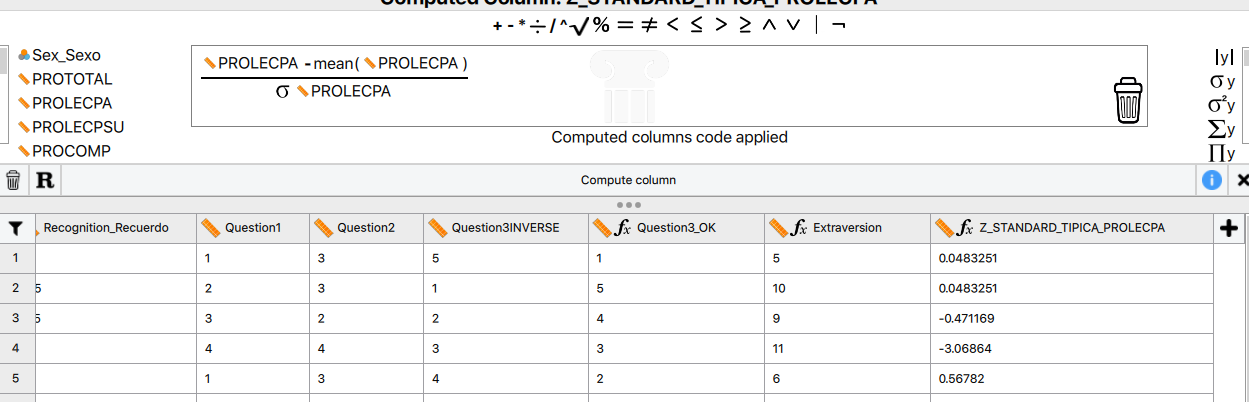
The answer is 6.2



*5. What is the z-score (standardized score) in PROLECPA for the person in the second row in the file. What does this z-score mean?*

We have is to calculate (Xi-M)/s as a new column





The value is .048, which means that the second person is 0.048 standard deviations above the average of the group. (That is, s/he is only slightly above the group average).